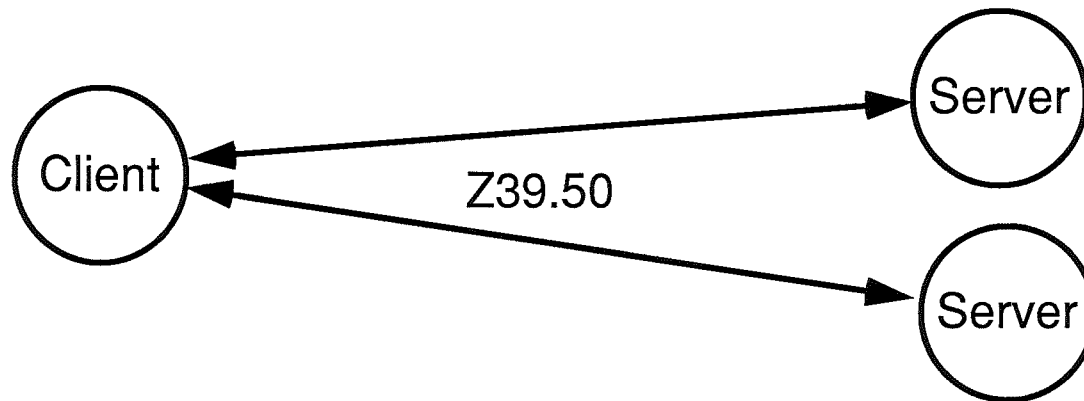"extra" slides

What is WAIS?

- WAIS Clients provide simple point-and-click interface

- Standard protocol connects clients to data servers

- WAIS Servers range from small to huge

- Improvements in servers and clients are independent

The WAIS Protocol *is* WAIS

```
                                    ( Server )
    ( Client )  ◄──────────────────►
              ◄──────    Z39.50
                    ──────────────►  ( Server )
```

- Supports any search syntax

- Supports sophisticated clients — puts intelligence in the user's hands

- Clients can run on any platform

- Multiple servers in a single search

- Retrieve any kind of data: text, graphics, video,...
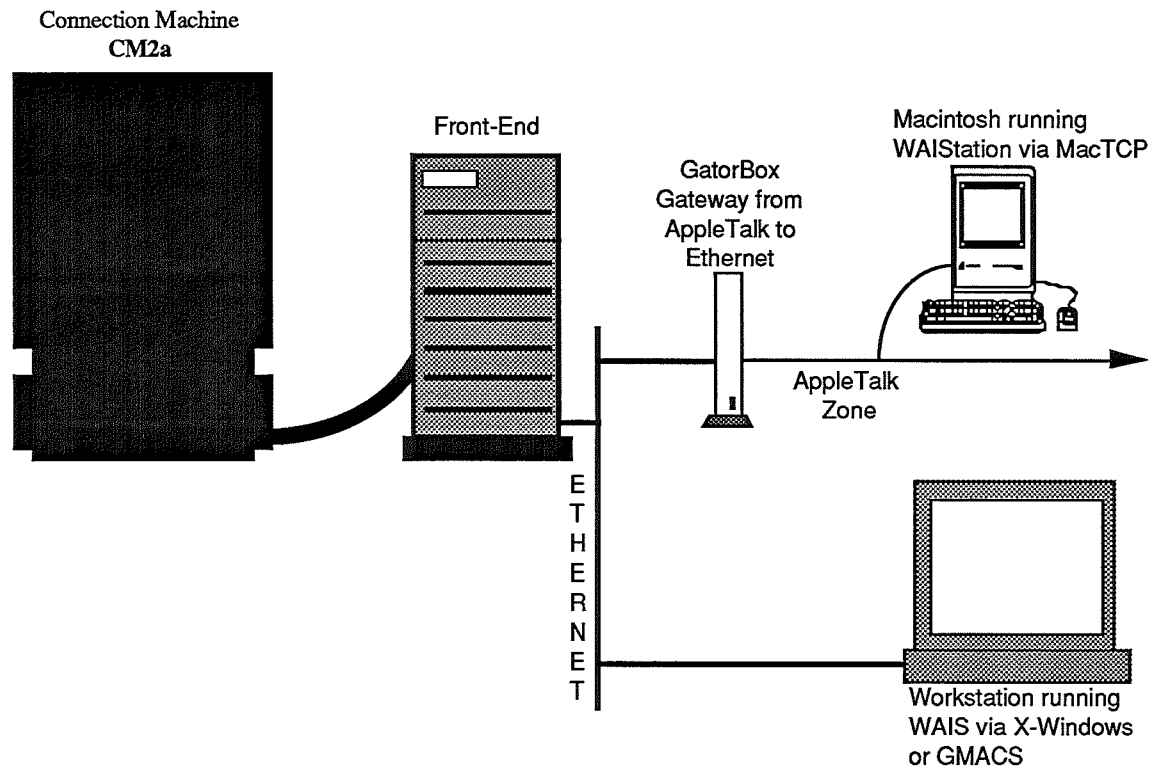
Levels of Information

- Personal files
  - Workgroup file server
    - Division database
      - Corporate/Organization database
        - Public databases

Goal: Access *all* levels from one interface

WAIS Hardware Components

Connection Machine
CM2a

Front-End

GatorBox
Gateway from
AppleTalk to
Ethernet

Macintosh running
WAIStation via MacTCP

AppleTalk
Zone

E
T
H
E
R
N
E
T

Workstation running
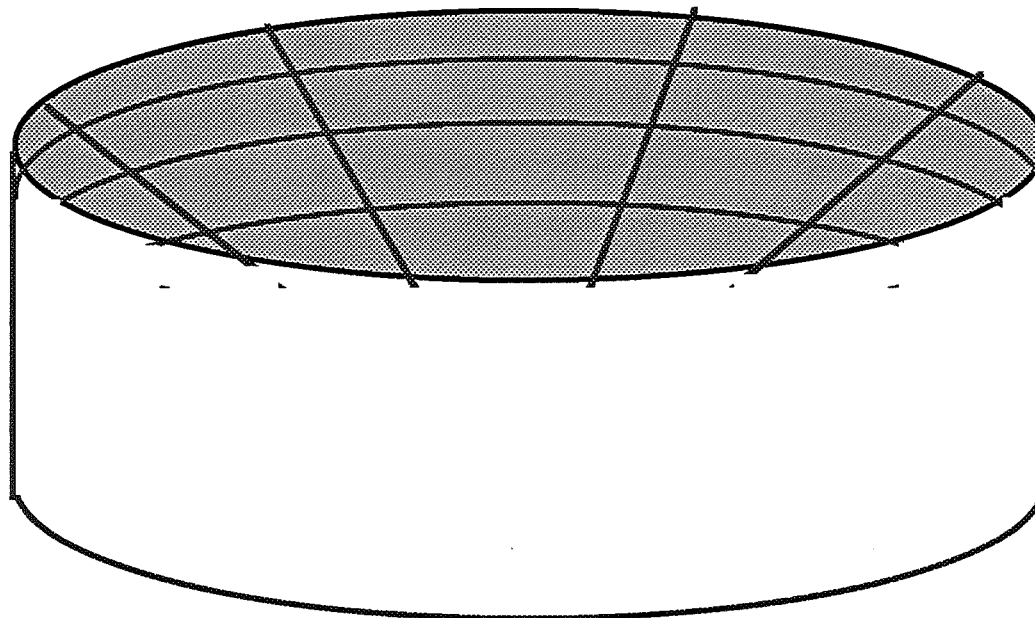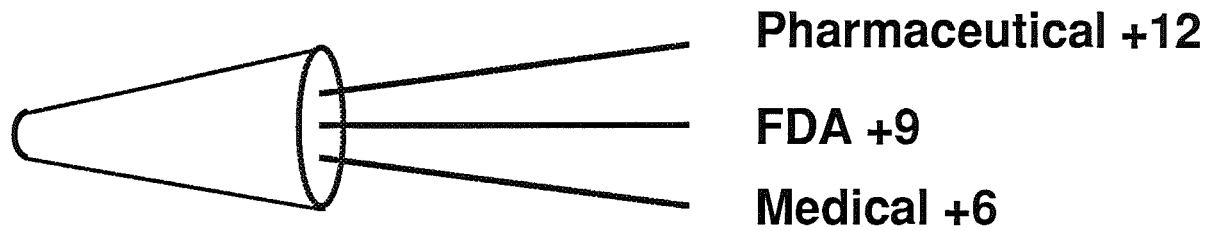WAIS via X-Windows
or GMACS

Why a Connection Machine?

- Bigger databases

- Interactive full-text search on gigabytes to terabytes

- More robust search techniques, e.g. relevance feedback, weighted terms

Data Parallelism:
Searching all the documents at once

Pharmaceutical +12

FDA +9

Medical +6

Stadium

Connection Machine Server

- 1-100GBytes (and getting bigger)

- Supports thousands of users

- Automatic Indexing

- Uses words and phrases in question to find appropriate documents with relevance feedback, weighted term

- Supports Boolean Queries

- Cost effective hardware alternative to mainframes

Document Retrieval Performance

- Current algorithm limits:

    ~2 GB with 512 MB CM-2

    ~8 GB with 2 GB CM-2

    ~25 GB with 8 GB CM-2

- High recall
- High precision } see Stanfill and Kahle
  Communications of the ACM
  December 1986

- << 1 sec. response

- Much larger DBs searchable with CM-5
 and inverted index algorithms: 100s to 1000s of Gigabytes

DowQuest

- An advanced information retrieval service offered by Dow Jones News/Retrieval since January 1989

- Simple and powerful *search by example* model.

- Prime the system with a few words to find an article you like.

- Search again using good article: "Give me more articles like that."

- The full text database of over 400 publications is examined and compared with the reference article.

- 16 top scoring "best fit" articles retrieved almost instantly.

- Process is repeated until you find just the information you want.

Boolean Search
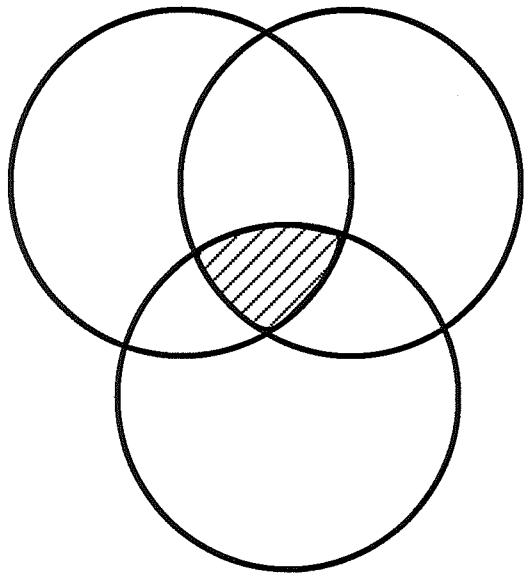
Conceptual Search

Retrieve documents
containing specific
combinations of
words
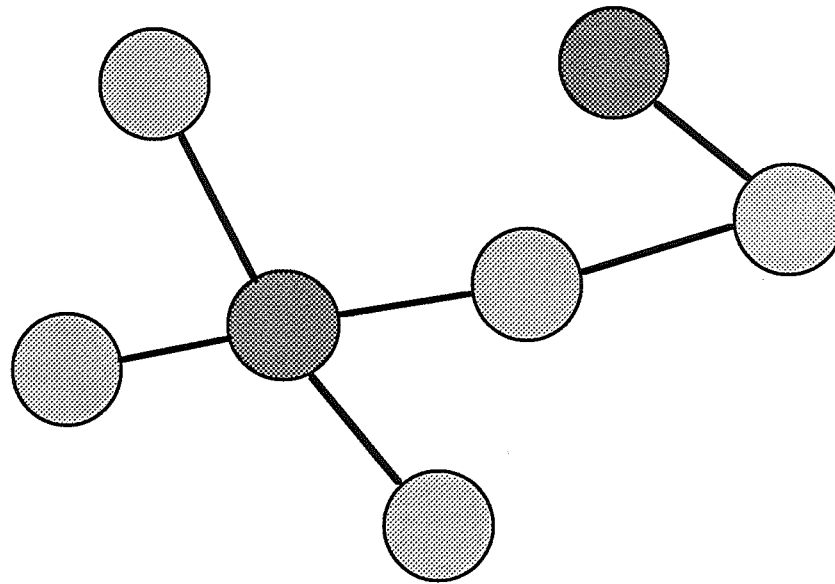
Explore a set of
documents
containing related
concepts

Boolean Query

Hard to Use:

Complex Syntax

> (Japanese OR Japan) AND
>
> (building OR buildings OR (Real AND Estate) AND
>
> (Manhattan OR (New AND York)

Poor Results:

The wrong information

No ranking of results

> Have you been paying attention?...
>
> Freer Finance: U.S. Regulators Move...
>
> REAL ESTATE: California Initiatives...
>
> First Boston Said To Agree on Sale Of...
>
> Exxon, Rockefeller Group to Sell Site...
>
> What's News--Business and Finance

Conceptual Search: Phase 1

Easy to Use:
No Syntax

Japanese buying real estate in mid-town manhattan

Options:

What do you
want to follow
up?

1. Time Acts to Cut Magazine Costs...
2. *First Boston Said To Agree on Sale...*
3. Have You Been Paying Attention?
4. *Exxon, Rockefeller Group to Sell Site...*
5. Hard Sell: Real Estate Developers...
6. What's News--Business and Finance...
7. Integrated Resources Buys Loft Building...

Conceptual Search: Phase 2

Relevance Feedback:

I like these; show me more

First Boston Said To Agree on Sale...

Exxon, Rockefeller Group to Sell Site...

Improved results:

Articles on related topics are found

Results are ranked

1. Bids for Exxon Building in New York...
2. Time Acts to Cut Magazine Costs...
3. Hard Sell: Real Estate Developers...
4. Time Inc. Sells Its 45% Interest...
5. Citicorp Unit Moves to Foreclose on...
6. Litigious Landlords: Legal Maneuvers

Results Improve with Query Size

Precision x recall @ 25% recall

Average performance over 13 reference sets

Typical Boolean Query

Typical Relevance Feedback Query

number of query terms.
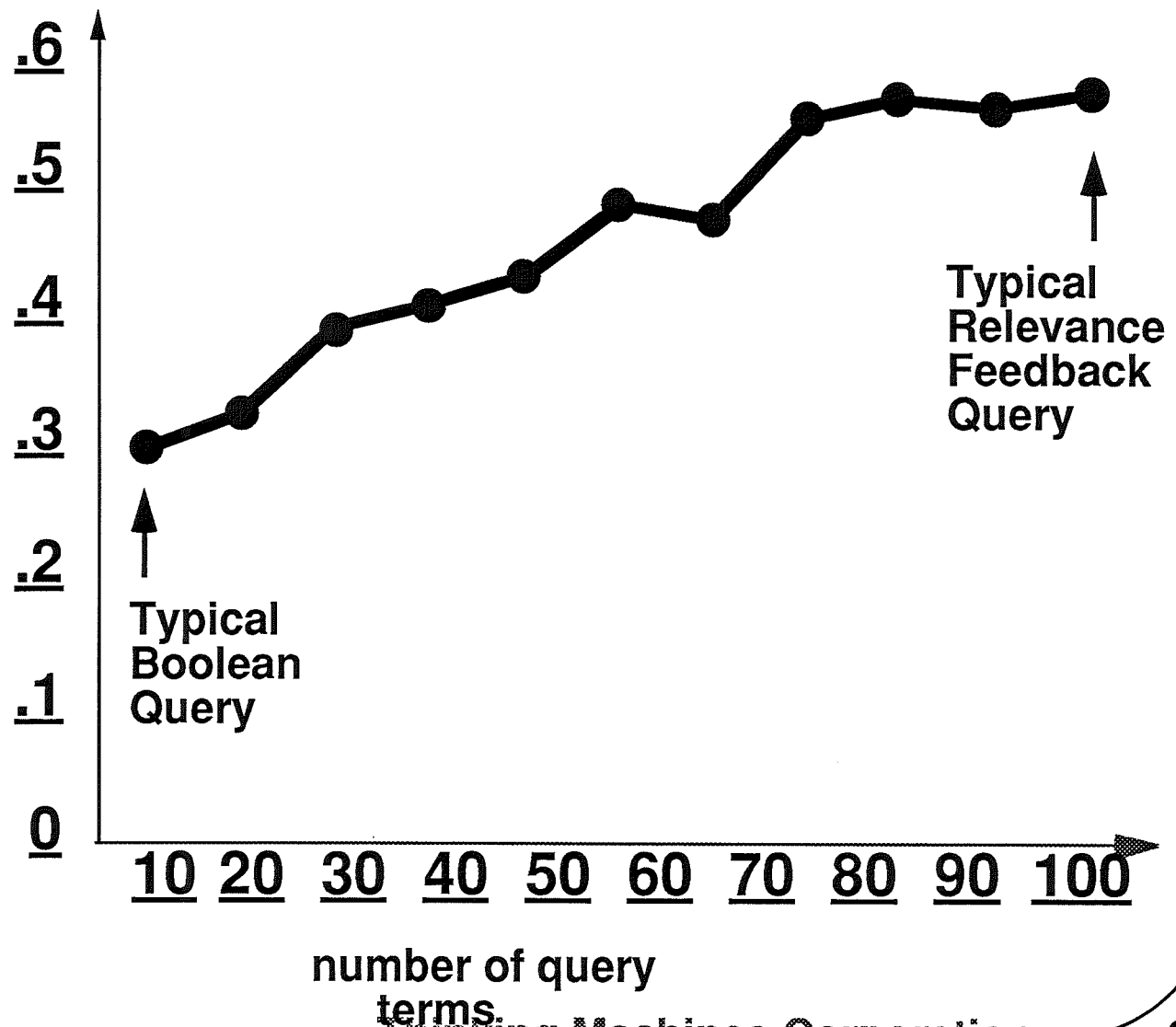
How Standard Protocol can Provide Security

- Users do not login to server, but search only through application layer protocol (Z39.50)

- Server controls access to data

- Network layers below application, or application layer handles authentication, encryption, billing

Public-Key Encryption
Fits the WAIS model

- Private data can be exchanged encrypted with public-key methods

- Once server has authenticated client, can encrypt data with client's public key

- Only client can decrypt, using private key

- Alternatively, user could buy key from server

- Layers below Application must handle encryption/decryption

WAIS

**Interface Design by
Thomas Erickson and Gitta Solomon**

**Graphic Design by
Ruth Ritter**

**Human Interface
Advanced Technology Group
Apple Computer, Inc.**

Thinking Machines Corporation

HD

The Only One

NEWS
8.23.90

Reporter

Search For   Motorola lawsuit

Search In   **Local**   **Remote**

☒ HD   ☒ DowQuest
☒ Company info   ☐ DowVision

Automate   ⦿ On   ◯ Off

⦿ Daily
◯ Weekly
◯ Other

[Search]

Trash

File Edit Reporter Newspaper Article Notebook

HD

The Only One

NEWS 8.23.90

Trash
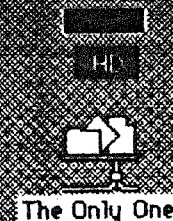
Specific Search

Results from [DowQuest ▼]

9 items
- ◆ Abreast of the Market: Stocks Fall on Investors'...
- ◆ What's News -- Business and Finance
- ◆ Technology: Motorola, Hitachi Reach a Draw...
- ◆ Motorola, Hitachi - Patent Suits -2-

[Get More] [Search Agent]

Preview

immediately between Hitachi and Motorola said
analyst Eli Sayegh of S.G. Warburg & Co. A resolution or
appeal is even more likely because of three additional
lawsuits that remain pending. Hitachi said one of the
future rulings should show that

[Open] [Save] [Add to Search]

Continue Search

contain:
(words) [Motorola lawsuit]

resemble:
(documents) ☐ Technology: Motorola, Hitachi Reach...

Search

[Now]
[Undo]

The General News

Index

Idle Reporters

Boycotts and high
   technology

Cold fusion

Australia and New
   Zealand

Space shuttle and
   NASA

Reporters

■ Motorola Lawsuit Hitachi

6 new items:

■ ◆◆◆◆

**Motorola, Hitachi Reach an
Agreement in Two Year Lawsuit**

Keyword. An instance lawsuit variable
is a special kind of variable used with
factories. A Motorola factory can
assign instance variables to specific
objects. Instance varibles contain a
unique set of values for each individual
object. The Hitachi methods of a fac-
tory can use the instance variables.

■ ◆◆◆

**Abreast of the Market: Stocks Fall on
Investor's Returns**

An instance stacks variable is a special
kind of variable used with factories. A
Motorola factory can as- sign instance
variables to specific objects. Instance
varibles contain a unique set of values
for each individual object. The Hitachi
methods of a factory can use the in.

■ New Human Interface Devices and Technologies

3 new items:

■ ◆◆◆◆

The Talking Glove

Although the agent in New Wave's
agent facility is a long way from agents
as envisioned by science fiction or the
Knowledge Navigator, it offers
functionality integrated into an
operating system.

■ ◆◆◆

Prototyping and User Testing

After the initial design interviews, we
conducted informal user tests. Test
subjects were Apple employees who
planned to attend the conference.

■ ◆◆

Animated Agents

Users were given the task of finding
certain types of presentations. For
example, we instructed them to find a

◀ 1 ▶

File   Edit   Reporter   Newspaper   Article   Notebook

High Technology Notebook

Organize by: date ▼

The Connection Machine and Local Area Networks

BYTE 8/31/89
Boring A. Reporter

There are two LAN-based information servers:
  1 Macintosh
  1 Small Connection Machine System.

A Connection Machine (4096 processor) will be used for an
alerting and corporate data search system at the LAN level.
This will get its data from the Macintosh X.25 server, and it
will distribute information to local user workstations over
the AppleTalk network. The front end that will be used for
this is a Sun4 configured to run a Connection Machine and
which can communicate with AppleTalk-based Macintoshes
through a GatorBox Ethernet/AppleTalk bridge.

The WAIS system will be the DowQuest information system.
The DowQuest system will be connected to the X.25
DowVision interface in the LAN Macintosh Server. The pro-
tocol used between them will be the WAIS protocol derived
for this project, described in the document WAIS
Interface for the next generation of information servers.

WAIS is a short term collaborative project between Apple,
Dow Jones, Peat Marwick and Thinking Machines. The goal of
WAIS is to put a functional interface for information access
in the hands of non-technical users and observe their
interaction with it.

- 1988
  - November
  - December
- 1989
  - January
  - February
  - March
  - April
  - May
  - June
  - July
  - **August**
  - September
  - October
  - November
  - December
- 1990
  - January
  - February
  - March
  - April

FIND